Handcrafted Features Fusion with Multilingual Models for Hope Speech Detection in Spanish and English

Rachana Nagaraju*,†, Hosahalli Lakshmaiah Shashirekha†

Department of Computer Science, Mangalore University, Mangalore, Karnataka, India

Abstract

Hope Speech detection in social media texts is essential for fostering supportive online communities and amplify positive and motivational content. However, identifying this content manually is cumbersome due to the ever growing amount of user-generated social media text. Further, the characteristics of user-generated text particularly when the information is subtle or masked by sarcasm makes it more challenging. To address these challenges, PolyHope at IberLEF 2025 shared task organized at IberLEF 2025 invites the research community to develop efficient models to detect Hope Speech tweets in Spanish and English languages. The shared task consists of two subtasks: i) Binary Hope Speech detection to classify the given tweet as 'Hope' or 'Not Hope' and ii) Multi-class Hope Speech detection to classify the given tweet as 'Generalized Hope', 'Realistic Hope', 'Unrealistic Hope', 'Sarcasm', or 'Not Hope', in Spanish and English. To tackle the challenges of detecting Hope Speech in English and Spanish, in this paper, we - team MUCS, describe the proposed multilingual models submitted to PolyHope at IberLEF 2025 shared task. Instead of developing seperate models for each language, we merge the train sets of English and Spanish, and build multilingual models using ensemble of Machine Learning (ML) classifiers (Logistic Regression (LR), Multinomial Naive Bayes (MNB), and Support Vector Machine (SVM)). Ensemble models are trained with fusion of selected handcrafted features (Term Frequency-Inverse Document Frequency (TF_IDF) vectors of words, character ngrams, and subwords) and their predictions are combined using hard and soft voting. The submitted models were evaluated by the organizers based on average macro F1 score and the best results obtained by our proposed ensemble models are: i) Binary classification in English using hard voting - macro precision of 0.80, macro recall of 0.79 and macro F1-score of 0.79, ii) Binary classification in Spanish using hard voting - macro precision of 0.78, macro recall of 0.78, and macro F1-score of 0.78, iii) multiclass classification in English using hard voting - macro precision of 0.64, macro recall of 0.63 and macro F1-score of 0.60, and iv) Multi-class classification in Spanish using soft voting - macro precision of 0.61, macro recall of 0.46, and macro F1-score of 0.49. These results highlight the effectiveness of soft voting in Multi-class scenario and demonstrate the robustness of Binary classification models in both languages. This research contributes to the development of inclusive language technologies, provides insights into the psychological and linguistic dimensions of hope, and supports the creation of more positive and resilient online environments.

Keywords

Hope Speech, Machine Learning, Multilingual models, Ensemble classifier, Byte Pair Encoding (BPE)

1. Introduction

Social media platforms have revolutionized the way people communicate, share ideas, and express emotions, creating a global platform for self-expression and community building. In addition to the benefits of connectivity and information sharing, social media house the proliferation of negative content in the form of hate speech, misinformation, etc., and also the positive content in the form Hope Speech. While the aim of negative content is to disturb social media ecosystem, Hope Speech that conveys optimism, encouragement, and positivity, plays a crucial role in fostering supportive and resilient online communities. Identifying and amplifying such Hope Speech can inspire individuals, promote mental well-being, and counteract the negative impacts of harmful content. The detection of Hope Speech in social media text is a multifaceted challenge due to the complexity of user-defined text, cultural nuances, and the subtle ways in which hope can be expressed, including sarcasm or indirect sentiments. Manual identification of Hope Speech is impractical given the sheer volume of

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

†These authors contributed equally.

rachananagaraju20@gmail.com (R. Nagaraju); hlsrekha@mangaloreuniversity.ac.in (H. L. Shashirekha)

0000-0002-9421-8566 (H. L. Shashirekha)

Table 1
Sample texts and their corresponding labels in English and Spanish

Language	Text	Binary Hope Speech	Multi-class Hope Speech
English	#USER# #USER# #USER# #USER# You expect a man that literally refers to himself as a god and is a blatantly obvious narcissist to be... humble?	Not Hope	Not Hope
English	Well the club season is over. We ended up tied for 44th place out of 130 in the Aspire division at the AAU National Championships. I had a great time! Played well.	Hope	Generalized Hope
English	On to Utica High School volleyball bright and early tomorrow. Go U!	Hope	Hope
English	RT #USER# #USER# Ojala nos ayuden a desenmascarar esta banda criminal	Hope	Realistic Hope
English	Find me a mass shooting in Utah where concealed carry on campus is legal.	Not Hope	Sarcasm
English	i wish my friends liked me more we could be smoking under this tree rn	Hope	Unrealistic Hope
Spanish	Outer Banks (temporada 3) la verdad que tenia muchas expectativas y me esperaba otra cosa. las escenas con el padre de john b las eliminaba no solo de mi	Hope	Generalized Hope
Spanish	RT #USER# #USER# Ojala nos ayuden a desenmascarar esta banda criminal	Hope	Realistic Hope
Spanish	Es como si estuviera enseñando historia a un montón de picaportes... Tuvimos una constitución antes de la actual. Se llamaba los Artículos de la Confederación.	Not Hope	Sarcasm
Spanish	Desearía no haber escuchado need y all of the girls	Not Hope	Not Hope
Spanish	Ojala pudiéramos saber en qué recuerdos nos tiene guardados una persona.	Hope	Unrealistic Hope

content generated daily on social media platforms like Twitter, Facebook, and Instagram. Therefore, automated approaches leveraging Natural Language Processing (NLP) and ML techniques have become indispensable for Hope Speech detection.

The PolyHope at IberLEF 2025: Optimism, Expectation or Sarcasm?[1] [1] shared task at IberLEF invites researchers to analyze and detect Hope Speech in social media texts, with emphasis on English and Spanish languages. The primary objective of this shared task is to classify tweets based on the expression of hope, even when the sentiments are subtly expressed or masked by complex language structures such as sarcasm. The task is divided into two main subtasks: Binary Hope Speech detection that involves classifying tweets into "Hope" or "Not Hope" and Multi-class Hope Speech detection that requires classifying tweets into one of the five categories: "Generalized Hope", "Realistic Hope", "Unrealistic Hope", "Not Hope", and "Sarcasm", in both languages. While Binary Hope Speech detection focuses on identifying both direct and subtle expressions of hope, especially when masked by complex language structures or implied sentiments, Multi-class Hope Speech detection involves distinguishing between different types of hope, and identifying sarcasm that mimics hopeful language. The sample texts and their corresponding labels in English and Spanish, from the given dataset are shown in Table 1.

To explore the strategies of detecting Hope Speech in Spanish and English, on social media platforms, in this paper, we - team MUCS, describe the multilingual models submitted to Polyhope at IberLEF 2025 shared task. This study aims to tackle the challenges of identifying Hope Speech in tweets by developing a multilingual framework that supports detection in both English and Spanish, thereby addressing the complexities of multilingual analysis. Multilingual models are designed to understand and process text in multiple languages simultaneously. In the era of globalization, the development of multilingual models has become increasingly crucial to bridge communication gaps between different languages

[1]https://www.codabench.org/competitions/5509/

and cultures. To facilitate the development of multilingual models, English and Spanish train sets are merged and considered as one train set for further processing. Our research builds on the existing NLP and ML techniques, incorporating: i) preprocessing techniques such as removing punctuation, converting emojis to text, and expanding contractions, ii) using TF-IDF of words, characters ngrams, and subwords, as features, and ii) employing an ensemble of ML classifiers - LR, MNB, and SVM, to enhance the performance of Hope Speech detection. The code to reproduce the proposed models is available in github[2]. This work contributes to the development of inclusive language technologies and provides valuable insights into the psychological and linguistic dimensions of hope. The findings will have broader implications for promoting positivity, inclusivity, and mental health in online environments, ultimately supporting the creation of more supportive and resilient digital communities.

The subsequent sections of this paper details the related works (Section 2), methodology (Section 3), experiments, results, and implications of our approach (Section 4), declaration on generative AI (Section 5) followed by conclusion and future works (Section 6).

2. Related Works

Hope Speech is believed to bring in positivity and helpful interactions in online spaces. Hence, social media platforms have been under the scrutiny of IT experts and social scientists for the detection of 'Hope Speech' in recent years. NLP and ML researchers have attempted to apply different approaches for Hope Speech detection and some of the notable works which contributed to the field are summarized below along their merits and weakness:

Lee et al. (2024) [2] achieved a significant milestone by integrating Graph NN with typical NLP techniques to better understand the relational patterns of Hope Speech on social media in English. They created a graph based on user interactions and text, which resulted in a macro F1 score of 0.87 on a popular Hope Speech dataset. This research showed that including structural and relational data enhances the performance and precision of tools used to detect Hope Speech. Johnson et al. (2024) [3], developed a transformer-based model using self-attention mechanisms to understand context in multilingual datasets made up of English, Spanish, and Hindi. The model achieved macro F1 scores of 0.88 for English, 0.85 for Spanish, and 0.82 for Hindi. Their work demonstrated that contextual embeddings and cross-lingual Transfer Learning (TL) have a strong influence on improving Hope Speech detection models helping them become both flexible and dependable.

Balouchzahi et al. (2023) [4] conducted a study to explore Hope Speech detection in English tweets by creating datasets to classify Hope Speech in both Binary and multi-class categories. These datasets were used to train ML, Deep Learning (DL), and TL frameworks, for detecting Hope Speech. Their findings illustrate that LR and CatBoost classifiers delivered better results than other methods by obtaining macro F1 scores of 0.80 and 0.79 in Binary classifications and 0.64 and 0.54 in multi-class tasks, respectively. This work provided benchmarks while proving the effectiveness of traditional ML models. Sidorov et al. (2023) [5] proposed a five layer Convolutional Neural Network (CNN) model to classify texts written in Spanish and English into "Hope" or "Non-Hope" categories. The results achieved macro F1 scores of 0.4974 for Spanish and 0.7238 for English. These findings show that DL algorithms have promise with multilingual data and intricate language characteristics.

Chakravarthi (2022) [6] worked on creating a Hope Speech dataset to address a research problem involving Equality, Diversity, and Inclusion (HopeEDI), in Hope Speech. This dataset meant to classify comments into two categories (Hope and Not Hope), contained data in English, Tamil, and Malayalam, created from YouTube comments. They built ML models including MNB, SVM, k-Nearest Neighbors (kNN), and Decision Tree (DT), using TF-IDF representations of word unigrams. Among these models, the DT classifier achieved the best results, with macro F1 scores of 0.46 for English and 0.56 for Malayalam. However, for Tamil, LR model produced the top score of 0.55. This research demonstrated that combining feature engineering with careful selection of ML algorithm could improve the detection effectiveness. Aggarwal et al. (2022) [7] worked on Hope Speech detection in English by relabeling the

[2] https://github.com/rachanabn20/MUCS-IberLEF2025

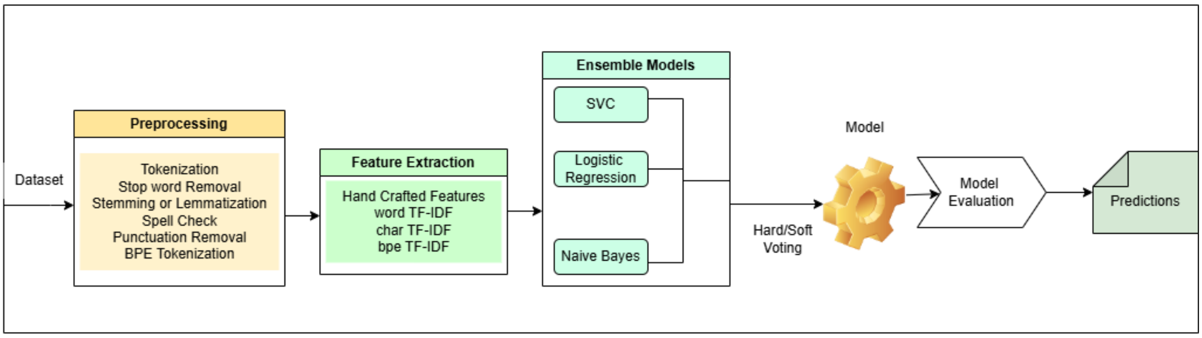

Figure 1: Framework of the proposed multilingual model for Hope Speech detection

data provided for "EACL-2021: Hope Speech Detection for Equality, Diversity, and Inclusion" shared task. They trained ML models (Naive Bayes, LR, and SVM) using TF-IDF of relabeled data and fine-tuned BERT with the relabeled data. Among these models, fine-tuned BERT model achieved the best results with a macro F1 score of 0.85, highlighting the superior performance of transformer-based models in capturing contextual information and nuanced sentiments. An ensemble of two DTs and a Random Forest classifier are ensembled with soft voting to select the most impactful features by Balouchzahi et al. (2022) [8] to enhance their model's performance. These features are used to trained a Keras Neural Network (NN) to detect Hope Speech. This approach achieved weighted F1 scores of 0.790 in Spanish and 0.870 in English.

Puranik et al. (2021) [9] explored transformer-based models to classify social media comments in English, Malayalam, and Tamil, as Hope or Not Hope. They experimented using BERT, ALBERT, DistilBERT, RoBERTa, CharacterBERT, ULMFiT, XLM-RoBERTa, MuRIL, mBERT and IndicBERT and the results illustrate that CharacterBERT and ULMFiT obtained solid F1 scores for English. To classify Malayalam, they found uncased mBERT paired with BiLSTM worked best. Their study highlighted the potential of transformer models to identify Hope Speech across various languages even with problems like uneven class sizes or mixed-language content.

These studies show different methods from traditional ML techniques to DL and TL approaches used to detect Hope Speech. Researchers have made progress, but challenges still exist as it is hard to manage multilingual data, recognize small hints of hope, and distinguish between hope and sarcasm. This research uses these earlier efforts to create a strong multilingual system to detect Hope Speech with the aim of building online spaces that are more positive and supportive.

3. Methodology

This study employs a systematic approach in building multilingual models to detect Hope Speech in tweets, focusing on English and Spanish languages. To facilitate the development of multilingual models, English and Spanish train sets are merged and considered as one train set for further processing. The methodology is structured into several key stages: text preprocessing, feature extraction, model training and model evaluation. The framework of the proposed multilingual model is shown in Figure 1 and the description of each stage highlighting the techniques and tools used to achieve accurate and reliable Hope Speech detection is given below:

3.1. Text Pre-processing

Effective pre-processing of text data is crucial for enhancing the performance of the learning models. The text is lower-cased and then the following pre-processing steps are applied to clean and normalize the text data:

- **Converting Emojis to Text**: Emojis are the pictorial representations of emotions. As they often convey sentiments and emotions, they are converted to their textual representations using the `emoji` [3] library.

- **Removing Punctuation**: Punctuation marks are removed from the text as they do not contribute significantly to identify Hope Speech.
  - Example: "hello, world!" → "hello world"
- **Expanding Contractions**: Contractions are expanded to their full forms using the *contractions*[4] library to standardize the text.
  - Example: "don't" → "do not"
- **Converting Numbers to Words**: Numeric values are converted to their word equivalents using the *num2words*[5] library to maintain uniformity in the text.
- **Removing Non-ASCII Characters**: Non-ASCII characters are removed to ensure compatibility with the models and to avoid encoding issues.
- **Removing Tags**: Hashtags and user mentions are removed as they do not contribute to the semantic meaning of the text.
- **Tokenization**: Text is split into meaningful tokens to facilitate further processing:
  - **Words**: The fundamental tokens in text are words. Hence, the raw text is split into individual words.
    * Example: "i love playing football" → ['i', 'love', 'playing', 'football']
  - **Character ngrams**: The contiguous sequences of *n* characters in a word forms the character ngrams. Character ngrams handles spelling variations and typos, useful for morphologically rich or low-resource languages, works well for short texts like tweets or usernames and helps to model prefixes/suffixes. Character ngrams in the range (1, 3) are extracted from the text.
    * Example: "playing" → ['p', 'l', 'a', 'y', 'i', 'n', 'g', 'pl', 'la', 'ay', 'yi', 'in', 'ng', 'pla', 'lay', 'ayi', 'yin', 'ing']

    Word and character ngrams tokens are obtained by using the NLTK [6] library.
  - **Subwords**: A sequence of characters such as character groups or morphemes, forms the subwords which help to handle out-of-vocabulary words, rare and compound words, in morphologically rich languages. Subwords are obtained using Byte Pair Encoding (BPE)[7].
    * Example: "playing" → ["play", "##ing"]
- **Removing Stopwords**: Common stopwords in English and Spanish are removed using the stopwords list available in NLTK library, as they do not carry significant semantic weights.
  - Example: "the quick brown fox" → "quick brown fox"
- **Lemmatization and Stemming**: English words are lemmatized (using WordNet Lemmatizer[8]) and/or stemmed (using Snowball Stemmer[9]) to reduce them to their root/stem forms, ensuring consistency in the text data.
  - Example: "better" → "good" (lemmatization)
  - Example: "jumps" → "jump" (stemming)

These preprocessing steps are applied to train, development and test sets to ensure consistency and improve the quality of the input data for feature extraction and model training.

[3] https://pypi.org/project/emoji/
[4] https://pypi.org/project/contractions/
[5] https://pypi.org/project/num2words/
[6] https://www.nltk.org/
[7] https://github.com/DolbyUUU/byte_pair_encoding_BPE_subword_tokenization_implementation_python
[8] https://www.nltk.org/api/nltk.stem.WordNetLemmatizer.html?highlight=wordnet
[9] https://www.nltk.org/api/nltk.stem.SnowballStemmer.html

Table 2
Hyperparameters and their values used in ML models

Classifier	Hyperparameter	Values
Logistic Regression	`max_iter`	1000
MultinomialNB	`alpha`	1.0
	`fit_prior`	True
SVC	`C`	1.0
	`kernel`	'rbf'
	`probability`	True

3.2. Feature Representation and Selection

Feature representation is a critical step in transforming raw text data into a format suitable for ML models. TF-IDF is a popular technique in NLP used to convert text into numerical features. TF-IDF vectors provide a normalized representation of text documents by highlighting the importance of words within the corpus. Tokenization resulted in 42,887 word unigrams, 14,525 character n-grams in the range of 1 to 3 and 25,229 subwords obtained from BPE, from the training set. *TfidfVectorizer*[10] from the *scikit-learn*[11] library is used to obtain vectors for these word unigrams, character n-grams, and subwords. Text data is high dimensional in nature and this increases the computational complexity. To reduce the computational complexity and consider the most informative features, chi-squared (χ^2)[12] statistical test is applied to each TF-IDF matrix to select top $1,000$ features. The selected features are combined to create a comprehensive feature set of size 3,000 to train the learning models.

3.3. Model Training

To improve the robustness and performance of the learning models, instead of using individual models, we have employed ensemble learning techniques with LR, MNB and SVM, as baselines with Hard and Soft voting.

- **Logistic Regression** is a linear classifier that models the relationship between features and the target variable using a logistic function
- **Multinomial Naive Bayes** is a probabilistic classifier based on Bayes' theorem suitable for text classification tasks
- **Support Vector Classifer** is a powerful classifier that finds the optimal hyperplane to separate data points in high-dimensional space

Soft voting combines the predicted probabilities of individual classifiers to make the final prediction whereas Hard voting combines the predicted labels of individual classifiers to make the final prediction. The hyperparameters and their values used in ML models/baselines are shown in Table 2. The ensemble models are trained on the selected feature sets to detect Hope Speech in English and Spanish.

4. Experiments and Results

This section details the experiments conducted to evaluate the performance of the proposed framework for Hope Speech detection in social media texts and the results obtained. Implementations are done in Google Colab using Python 3.11.12 and scikit-learn 1.6.1[13] library for ML models. Various experiments were conducted to train different combination of baselines for ensembling to detect Hope Speech in social media texts and the models that performed well on development sets (Dev) are used to predict the class labels of test sets for both Binary and Multi-class categories. The performance of the models is

[10]https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
[11]https://scikit-learn.org/stable/
[12]https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html
[13]https://scikit-learn.org/stable/

Table 3
Statistics of the datasets

Language	Subtask	Category	Train set	Dev set
Spanish	a) Binary Hope Speech detection	Not Hope	5,316	1,926
		Hope	5,927	2,162
	b) Multi-class Hope Speech detection	Not Hope	5,383	1,958
		Generalized Hope	2,754	1,001
		Unrealistic Hope	1,300	473
		Realistic Hope	1,113	405
		Sarcasm	693	251
English	a) Binary Hope Speech detection	Not Hope	2,807	1,003
		Hope	2,426	899
	b) Multi-class Hope Speech detection	Not Hope	2,245	816
		Generalized Hope	1,284	467
		Unrealistic Hope	472	171
		Realistic Hope	540	196
		Sarcasm	692	252

Table 4
Performances of the proposed multilingual models in Binary Hope Speech detection

Language	Voting	Dev Set			Test Set		
		Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
English	Hard Voting	0.79	0.79	0.78	0.80	0.79	**0.79**
English	Soft Voting	0.78	0.78	0.78	0.80	0.79	**0.79**
Spanish	Hard Voting	0.78	0.77	0.77	0.78	0.78	**0.78**
Spanish	Soft Voting	0.77	0.77	0.77	0.78	0.78	**0.78**

Table 5
Performances of the proposed multilingual models in Multi-class Hope Speech detection

Language	Voting	Dev Set			Test Set		
		Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
English	Hard Voting	0.62	0.61	0.58	0.64	0.63	**0.60**
English	Soft Voting	0.63	0.63	0.61	0.65	0.52	**0.55**
Spanish	Hard Voting	0.60	0.63	0.58	0.62	0.44	**0.47**
Spanish	Soft Voting	0.61	0.63	0.59	0.61	0.46	**0.49**

evaluated based on macro F1 score as it provides a balanced measure of precision and recall across the classes.

4.1. Dataset Description

The statistics of the Train and Development (Dev) sets provided by the organizers of the shared task are shown in Table 3. They also provided Test sets of size 2,380 for English and 5,111 for Spanish. Even though the given Train datasets are imbalanced, we didn't put any efforts to balance them. These datasets have been extensively used in several previous studies for Hope Speech detection [4, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19].

4.2. Results

The performances of the proposed ensemble models on the Development and Test sets of English and Spanish languages for Binary and Multi-class Hope Speech detection are shown in Tables 4 and 5 respectively. In Binary Hope Speech detection task, using Hard and Soft voting the models achieved

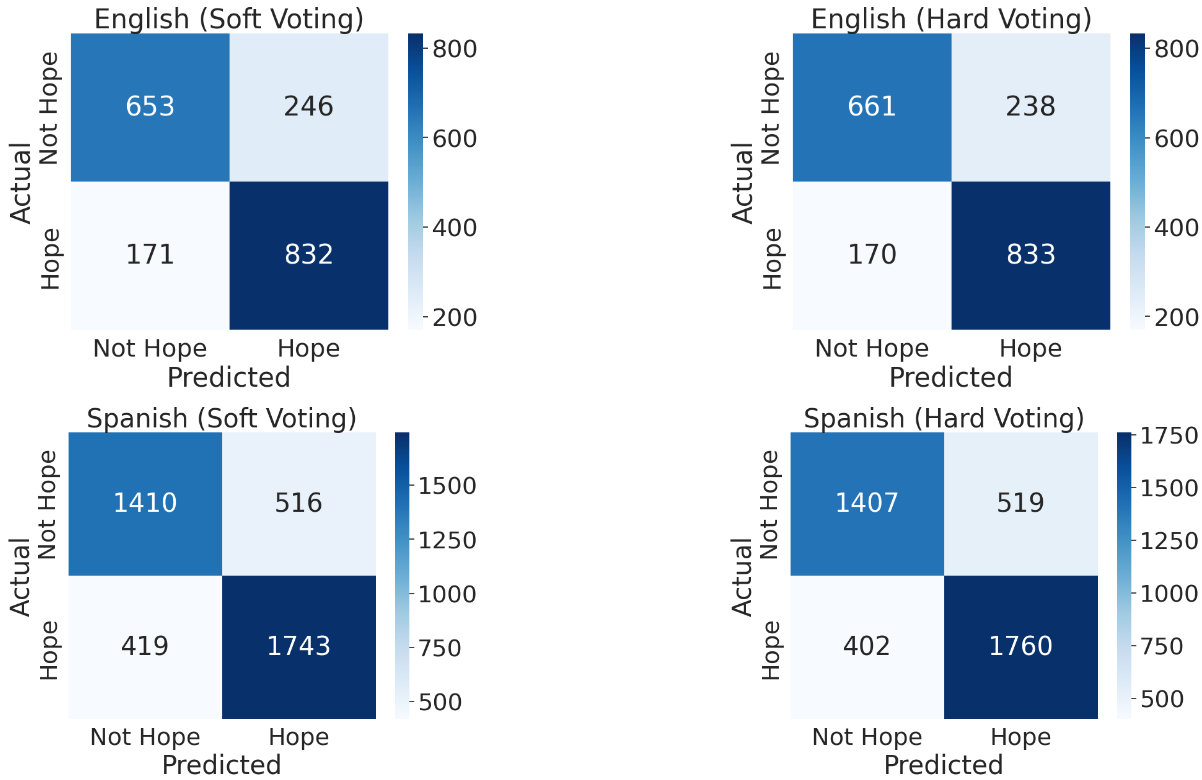

Figure 2: Confusion matrix for Binary classification in English and Spanish Development sets

identical performance with Macro F1 scores of 0.79 and 0.78 for English and Spanish, respectively, on Test sets. However, in Multi-class Hope Speech detection task, the performance dropped, with macro F1 score of 0.60 and 0.55 using Hard and Soft voting, respectively, for English, and 0.47 and 0.49 using Hard and Soft voting, respectively for Spanish, for Test sets. Overall, models performed better in the Binary classification task, with a noticeable performance drop in the Multi-class classification task, particularly for Spanish.

4.3. Error Analysis

Analyzing errors is the key to enhance model performance and reliability. As the Test data with labels is not available, error analysis is based on the results of the Development sets. Figures 2 and 3 present the confusion matrices for Binary and Multi-class classification, respectively, using hard and soft voting for English and Spanish. Based on these confusion matrices, we performed an in-depth error analysis to assess the performance of the classification models.

For Binary classification in English, soft voting model mislabeled "Hope" much more than "Not Hope" (17.04% samples are misclassified as "Not Hope" and 27.36% samples are classified as "Hope"). The hard voting model showed a similar trend (16.94% samples are misclassified as "Not Hope" and 26.47% samples are classified as "Hope"). For Binary classification in Spanish, our models did pretty well at identifying "Not Hope" but had trouble with "Hope". Using soft voting, 19.38% samples are misclassified as "Not Hope" and 26.79% samples are classified as "Hope". Similarly, with hard voting, 18.59% samples are misclassified as "Not Hope" and 26.94% samples are classified as "Hope".

The multiclass classification task brought its own challenges. For English dataset, "Generalized Hope" often ended up being misclassified as "Not Hope," with 156 misclassifications under soft voting and 152 under hard voting. "Realistic Hope" faced significant issues too, frequently confused with "Not Hope" and "Generalized Hope," with 35 and 31 instances misclassified under soft and hard voting, respectively. Misclassifications were also common for "Sarcasm" and "Unrealistic Hope," primarily labeled as "Not Hope," with "Sarcasm" seeing 90 instances misclassified under soft voting and 134 under hard voting.

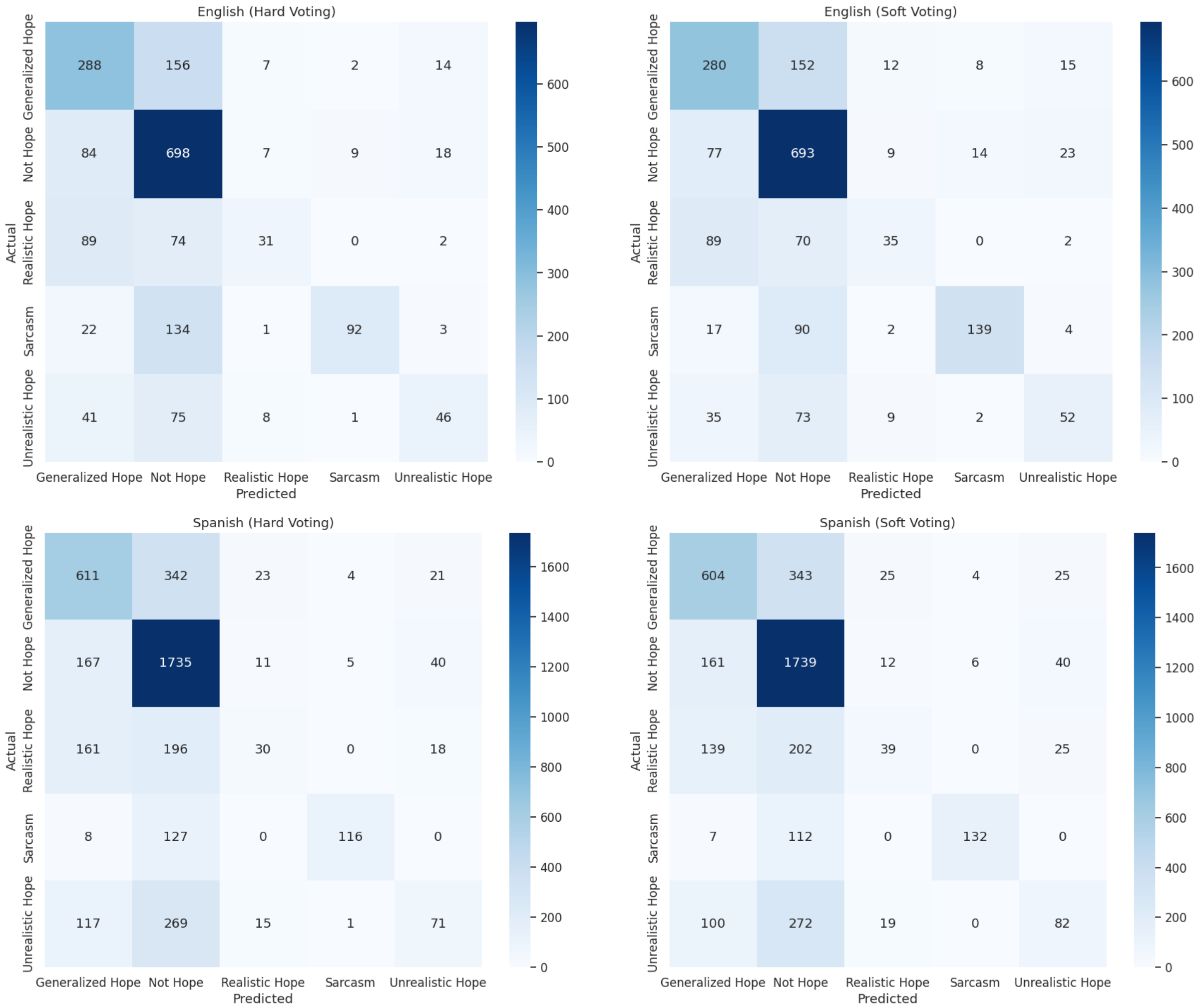

Figure 3: Confusion matrix for Multi-class classification in English and Spanish Development sets

Table 6
Sample Misclassified Texts in Binary Classification from Development Set

Language	True Label	Predicted Label	Text
Spanish	Hope	Not Hope	Las personas que menos espero que estén son las más incondicionales en esta etapa de miedo de mi vida. A esas personas gracias áÁî,đđđđ
Spanish	Not Hope	Hope	á¦Circulasica. El sueÁo de Milikiá¦e ofrecerÁ esta semana sus funciones en el Auditorio Ciudad de LeÁn https://t.co/l8F0KN6cEt
English	Hope	Not Hope	also that jacket he wore yesterday áááááám gonna rip my eyes out :/ hey guys! áááááám so happy to announce that á ááááám officially brain dead
English	Not Hope	Hope	Job Alert: Cfo (#Yonkers, New York) Aspire Partners, LLC #Job #Finance #CostReduction #AccountsPayable #RelationshipManagement #Sales #TrainingandDevelopment #RelationshipBuilding #Management

Likewise, "Unrealistic Hope" experienced notable misclassification challenges, with 73 misclassifications under soft voting and 75 under hard voting. In the Spanish dataset, "Generalized Hope" was regularly misclassified as "Not Hope," with 343 instances under soft voting and 342 under hard voting. "Realistic

Table 7
Sample Misclassified Texts for Multi-class Classification from Development Set

Language	True Label	Predicted Label	Text
English	Generalized Hope	Not Hope	And youâ€™d come to the obvious logical solution: do whatever it takes to get those 2 seats. The goal is hard but not monumental â€” definitely do-able!! We have excellent prospects in Val Demmings in FL, Cheri Beasley in NC, John Fetterman in FL!! There is reason for hope here.
English	Realistic Hope	Generalized Hope	#USER# Iâ€™m hoping his type helps quiet our world down a bit. He teaches me that you can be very bossy without any words. He teaches me to read energy. He really is extraordinary.
English	Unrealistic Hope	Not Hope	I aspire to be his sugar mommy...I know chance of it happening is like ...Never(i am not that much delusional)...But I am gonna die trying!
English	Sarcasm	Not Hope	also that jacket he wore yesterday â€¦â€¦â€¦â€¦â€¦â€¦â€¦m gonna rip my eyes out :/
English	Not Hope	Unrealistic Hope	#USER# i wish i could help but iâ€™m not that good at making flags and even if i was making flags for an identity thatâ€™s not yours is weird imo
Spanish	Realistic Hope	Not Hope	Estoy esperando que nazca mi sobri solo por el hecho de que ese dÃa no trabajo ð
Spanish	Generalized Hope	Not Hope	Lo emocionada que estoy con los oscars la semana que viene (creo que es) ya sÃ©Â© que Jessica no estÃ¡ nominada ni nada pero me da ilusiÃ³n verla en la alfombra roja y presentar ððð, IntentarÃ© irme a dormir tarde ese dÃa para verla.
Spanish	Realistic Hope	Not Hope	Estoy deseando ver esta noche el programa para escuchar todas las burradas contra el barcelona, las protestas de SÃ¡nchez que es un aburrimiento oÃrlo y por Ãºltimo las imÃ¡genes del Ãjrbito Munuera diciendo que la tarjeta aVinicius era roja ChiringuitoDeMega
Spanish	Not Hope	Generalized Hope	Que hechizo tengo que hacer para que se cumpla esto? A quien le rezo? #URL#
Spanish	Unrealistic Hope	Not Hope	Quedate con alguien que te mire, con el anhelo que Amelia mira a Luisita. #URL#
Spanish	Sarcasm	Not Hope	Me encanta cuando los seguidores de mitos aleatorios hablan sobre pensamiento crÃ¡tico :v

Hope" also had a high number of misclassifications as "Not Hope," with 202 instances under soft voting and 196 under hard voting. Both "Sarcasm" and "Unrealistic Hope" faced considerable challenges, mainly being misclassified as "Not Hope," with "Sarcasm" seeing 112 misclassifications under soft voting and 127 under hard voting. "Unrealistic Hope" had 272 instances misclassified under soft voting and 269 under hard voting.

Several factors have played role in these misclassifications. The datasets contained linguistic complexities, like slang, idiomatic phrases, and specialized terminology, which created hurdles for the models. Contextual misunderstandings further exacerbated the classification challenges, as the models struggled with the subtleties in which certain terms are used. Class imbalances in the datasets also affected the ability of the models to accurately predict certain classes. Additionally, the features selected to represent the data and the intrinsic limitations of the model architectures also have contributed to the misclassifcation. This analysis highlights the complex challenges of accurately classifying text data, especially across different languages and tasks, and points to the need for ongoing refinement in model training and feature engineering to boost classification performance.

5. Declaration on Generative AI

To prepare this work, we used generative AI tools to analyze data, review related works, create code, and edit draft. These tools helped to spot trends in data and sum-up prior research. We reviewed the text and checked all AI-generated content to ensure accuracy. Our critical thinking and contributions shaped the final content and results. AI supported the process but did not replace human judgment.

6. Conclusion and Future Work

In this paper, we - team MUCS, describe the multilingual ensemble models submitted to PolyHope at IberLEF 2025 shared task organized at IberLEF 2025 for Hope Speech detection in Spanish and English. Using feature fusion of TF-IDF of selected words, character ngrams and subwords, ensemble models with LR, MNB and SVM baselines, are trained to predict the class labels with Hard and Soft voting for the test data, in both Binary and Multi-class Hope Speech detection. Our proposed models achieved notable results, macro F1-scores of 0.78 and 0.79 for Binary Hope Speech detection in Spanish and English, respectively, using Hard voting. For Multi-class Hope Speech detection, the models attained macro F1-scores of 0.49 and 0.60 for Spanish and English, respectively, using Hard voting. This research contributes to the advancement of inclusive language technologies, offering insights into the psychological and linguistic dimensions of hope, and fostering the creation of more positive and resilient online environments. Improving the generalization of models in Multi-class settings and exploring additional features or advanced techniques to better capture the nuances of Hope Speech across languages, will be explored in future.

References

[1] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and Other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

[2] D. Lee, E. Wang, F. Kim, Integrating Graph Neural Networks with NLP for Enhanced Hope Speech Detection, International Journal of Social Media Analysis 8 (2024) 112–130.

[3] A. Johnson, B. Smith, C. Doe, Advances in Hope Speech Detection using Transformer-based Models, Journal of Natural Language Processing 12 (2024) 45–67.

[4] F. Balouchzahi, G. Sidorov, A. Gelbukh, PolyHope: Two-level Hope Speech Detection from Tweets, in: Expert Systems with Applications, Elsevier, 2023, pp. 123–456.

[5] M. Shahiki-Tash, J. Armenta-Segura, O. Kolesnikova, G. Sidorov, A. Gelbukh, Lidoma at Hope2023IberLEF: Hope Speech Detection using Lexical Features and Convolutional Neural Networks, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS. org, 2023.

[6] B. R. Chakravarthi, Hope Speech Detection in YouTube Comments, Social Network Analysis and Mining 12 (2022) 75.

[7] P. Aggarwal, P. Chandana, J. Nemade, S. Sharma, S. Saumya, S. Biradar, Hope Speech Detection on Social Media Platforms, 2022. arXiv:2212.07424.

[8] F. Balouchzahi, S. Butt, G. Sidorov, A. Gelbukh, CIC@LT-EDI-ACL2022: Are Transformers the only Hope? Hope Speech Detection for Spanish and English Comments, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 206–211. doi:10.18653/v1/2022.ltedi-1.28.

[9] K. Puranik, A. Hande, R. Priyadharshini, S. Thavareesan, B. R. Chakravarthi, IIITT@ LT-EDI-EACL2021-Hope Speech Detection: There is Always Hope in Transformers, arXiv preprint arXiv:2104.09066 (2021).

[10] S. Butt, F. Balouchzahi, M. Amjad, S. Jiménez-Zafra, H. Ceballos, G. Sidorov, Overview of Polyhope at IberLEF 2025: Optimism, Expectation or Sarcasm?, Procesamiento del Lenguaje Natural (2025).

[11] S. Butt, F. Balouchzahi, A. Amjad, M. Amjad, H. Ceballos, S. Jiménez-Zafra, Optimism, Expectation, or Sarcasm? Multi-class Hope Speech Detection in Spanish and English (2025). URL: https://doi.org/10.13140/RG.2.2.19761.90724. doi:10.13140/RG.2.2.19761.90724.

[12] G. Sidorov, F. Balouchzahi, S. Butt, A. Gelbukh, Regret and Hope on Transformers: An Analysis of Transformers on Regret and Hope Speech Detection Datasets, Applied Sciences 13 (2023) 3983.

[13] D. García-Baena, F. Balouchzahi, S. Butt, M. García-Cumbreras, A. Tonja, J. García-Díaz, et al., Overview of Hope at IberLEF 2024: Approaching Hope Speech Detection in Social Media from Two Perspectives, for Equality, Diversity and Inclusion and as Expectations, Procesamiento del lenguaje natural 73 (2024) 407–419.

[14] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope Speech Detection in Spanish: The LGBT Case, Language Resources and Evaluation 57 (2023) 1487–1514.

[15] S. Jiménez-Zafra, M. Garcia-Cumbreras, D. García-Baena, J. Garcia-Díaz, B. Chakravarthi, R. Valencia-García, L. Ureña-López, Overview of Hope at IberLEF 2023: Multilingual Hope Speech Detection, Procesamiento del Lenguaje Natural 71 (2023) 371–381.

[16] B. R. Chakravarthi, HopeEDI: A Multilingual Hope Speech Detection Dataset for Equality, Diversity, and Inclusion, in: Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, 2020, pp. 41–53.

[17] B. Chakravarthi, V. Muralidaran, R. Priyadharshini, S. Cn, J. McCrae, M. García, S. Jiménez-Zafra, R. Valencia-García, et al., Overview of the Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, ACL, 2022, pp. 378–388. doi:10.18653/v1/2022.ltedi-1.58.

[18] G. Sidorov, F. Balouchzahi, L. Ramos, H. Gómez-Adorno, A. Gelbukh, Mind-hope: Multilingual Identification of Nuanced Dimensions of HOPE (2024).

[19] F. Balouchzahi, S. Butt, M. Amjad, G. Sidorov, A. Gelbukh, Urduhope: Analysis of Hope and Hopelessness in Urdu Texts, Knowledge-Based Systems 308 (2025) 112746.